WELCOME TO THE 2022 NDACAN SUMMER TRAINING SERIES!

- July 27, 2022 Structural Equation Modeling Workshop
- The session will begin at 12pm EST.
- Please submit questions to the Q&A box.
- This session is being recorded.

NDACAN SUMMER TRAINING SERIES

National Data Archive on Child Abuse and Neglect Cornell University & Duke University





STRUCTURAL EQUATION MODELING

NDACAN SUMMER TRAINING SERIES SCHEDULE

- July 6, 2022 Introduction to NDACAN and Data Management Strategies
- July 13, 2022 Administrative Data (NCANDS, AFCARS) and Linking
- July 20, 2022 Linking NDACAN Data with External Products
- July 27, 2022 Structural Equation Modeling Workshop
- August 3, 2022 Propensity Score Matching Workshop
- August 10,2022 Studying Racial Disparities using NDACAN Data

SESSION AGENDA

- What is SEM
- SEM process
- Example

WHAT IS STRUCTURAL EQUATION MODELING?

WHAT IS SEM?

- Structural equation modeling (SEM) encompasses a broad array of statistical techniques and frameworks
 - Family of related methods rather than single technique
- Tries to help explain relationships between observed/measured variables and latent variables
 - Latent variables are variables that can't be directly measured but have an underlying impact on the response and other variables, e.g. intelligence, religiosity, quality of life
- Combination of path analysis, factor analysis, and multiple regression analysis
 - If no latent variables, then analogous to path analysis

WHAT IS SEM?

- The following relationships are possible in SEM:
 - observed to observed variables (e.g. regression)
 - latent to observed variables (e.g. confirmatory factor analysis)
 - latent to latent variables (e.g. structural regression)
- Encompasses both measurement and structural models
 - Measurement model relates observed to latent variables
 - The paths that run from latent variables to measured variables are factor loadings expressing the strength of the relationship between an indicator and a latent variables
 - Analogous to factor analysis
 - Structural model relates latent to latent variables
 - The paths between the LVs express the strength and direction of the relationships as regression coefficients.

WHY SEM?

- More flexibility than doing factor analysis and/or regression analysis piecewise
- Well-suited for causal analysis
- Can handle multicollinearity

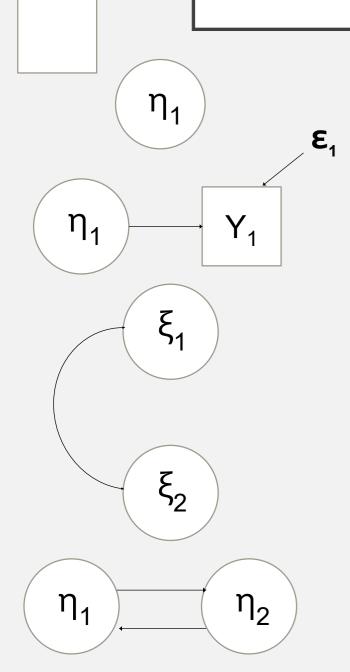
PATH DIAGRAMS

- Visual representation of system of simultaneous equations
- Each shape, line, and arrow in a path diagram has meaning

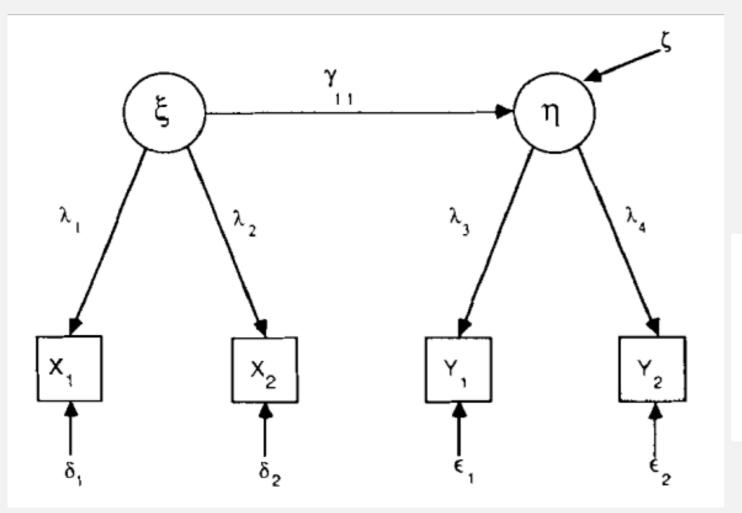
PATH DIAGRAMS

Primary Symbols Used in Path Analysis

- Rectangular or square box signifies an observed or manifest variable.
- Circle or ellipse signifies an unobserved or latent variable.
- Unenclosed variable signifies a disturbance term (error in either equation or measurement).
- Straight arrow signifies assumption that variable at base of arrow "causes" variable at head of arrow.
- Curved two-headed arrow signifies unanalyzed association between two variables
- Two straight single-headed arrows connecting two variables signifies feedback relation or reciprocal causation
- *Figures and text from K. Bollen, Structural Equations with Latent Variables,
 Table 2.6



PATH DIAGRAMS



• The path diagram in Figure 2.4 is equivalent to the following simultaneous system of equations:

$$\eta = \gamma_{11}\xi + \zeta$$

$$x_1 = \lambda_1\xi + \delta_1, \qquad x_2 = \lambda_2\xi + \delta_2$$

$$y_1 = \lambda_3\eta + \epsilon_1, \qquad y_2 = \lambda_4\eta + \epsilon_2$$

Figure 2.4 An Example of a Path Diagram

CONSIDERATIONS AND ASSUMPTIONS

- Considerations:
 - Outliers
 - Missing data
 - Degrees of freedom (e.g. sample size and number of unknowns to estimate)
- Assumptions:
 - SEM: multivariate normality
 - Model specified correctly

SEM PROCESS

STEPS TO SEM ANALYSIS

Specify model

 Create path diagram with measurement and structural model of interest

2. Evaluate model

 Check measures and distributional assumptions of each variable of interest

3. Fit model and assess fit

 Use likelihood ratio Chi-squared test (want p-value > 0.05),
 AIC, BIC, R², root mean square error of approximation (RMSEA; want RMSEA < 0.06 or p-value > 0.05)

PROGRAMMING

- Stata
 - sem
- R
 - lavaan
 - sem
- SAS
 - PROC CALIS

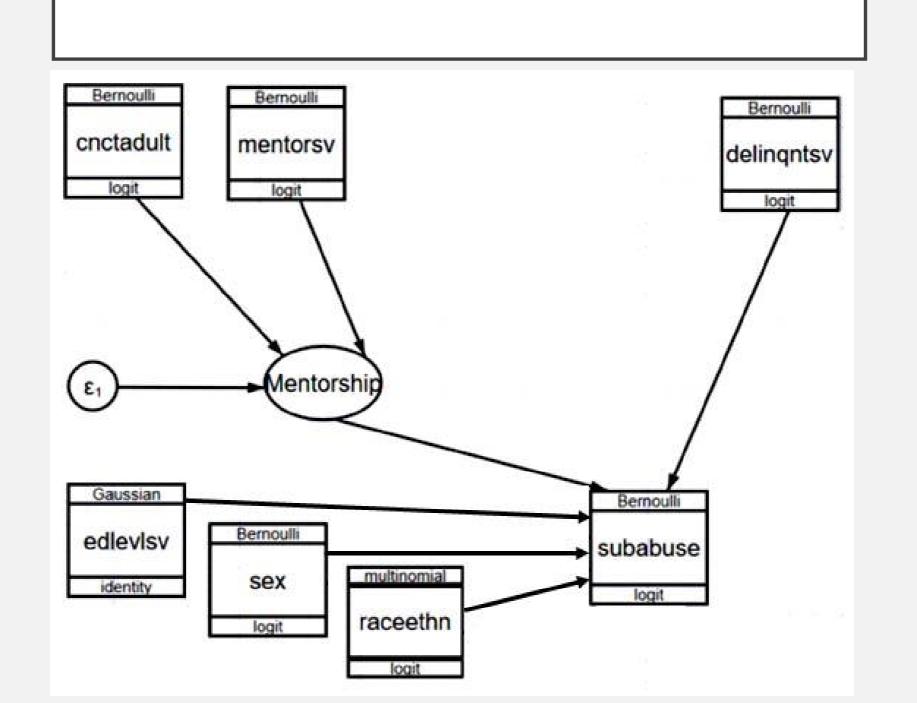
EXAMPLE

EXAMPLE USING LINKED NYTD SERVICES AND OUTCOMES

- Link 2014 cohort NYTD service file with outcomes file
 - The Services component (this dataset) contains crosssectional information on the services provided by states under Chafee Program and the youth who receive those services.
 - The Outcomes component contains the results of surveys conducted with youth to examine certain well-being, financial, and educational outcomes as they get older.
- Summarize to one row per child,
 - Services ever received
 - Last outcomes known wave 3

EXAMPLE USING LINKED NYTD SERVICES AND OUTCOMES

- Want to understand relationship of services received while aging out of foster care and having a substance abuse problem
- Observed variables
 - Delinquency
 - Education level
 - Sex
 - Race/ethnicity
- Latent variables
 - Measure of mentorship/guidance



REFERENCES

- Kline, Rex B. Principles and practice of structural equation modeling. Guilford publications, 2015.
- Bollen, Kenneth A. Structural equations with latent variables. Vol. 210. John Wiley & Sons, 1989.
- [SEM] structural equation modeling stata. Retrieved May 16, 2022, from https://www.stata.com/manuals13/sem.pdf
- Jerry Bean, C. of S.W. (2021, February 26). *Using R for social work research*. 5 Structural Equation Modeling. Retrieved May 16, 2022, from https://bookdown.org/bean_jerry/using_r_for_social_work_research/struct ural-equation-modeling.html
- Acock, Alan C. "Discovering structural equation modeling using Stata." Stata Press Books (2013).
- Introduction to Structural Equation Modeling (SEM) in R with lavaan. UCLA: Statistical Consulting Group. Retrieved May 16, 2022, from https://stats.oarc.ucla.edu/r/seminars/rsem/

CODE IN STATA

import delimited "C:\Users\ss3376.NDACAN\Box\NDACAN\Presentations\Summer Series 2022\S4 SEM\s4_nytd_ex.csv", stringcols(I) clear

logit subabuse raceethn delinqntsv edlevlsv mentor homeless sex

QUESTIONS?

SARAH SERNAKER STATISTICIAN

SARAH.SERNAKER@DUKE.EDU

NEXT WEEK...

•Topic:

Propensity score matching

•Presenter:

Alex Roehrkasse

•Date:

•August 3rd, 2022